Learning Continual Compatible Representation for Re-indexing Free Lifelong Person Re-identification

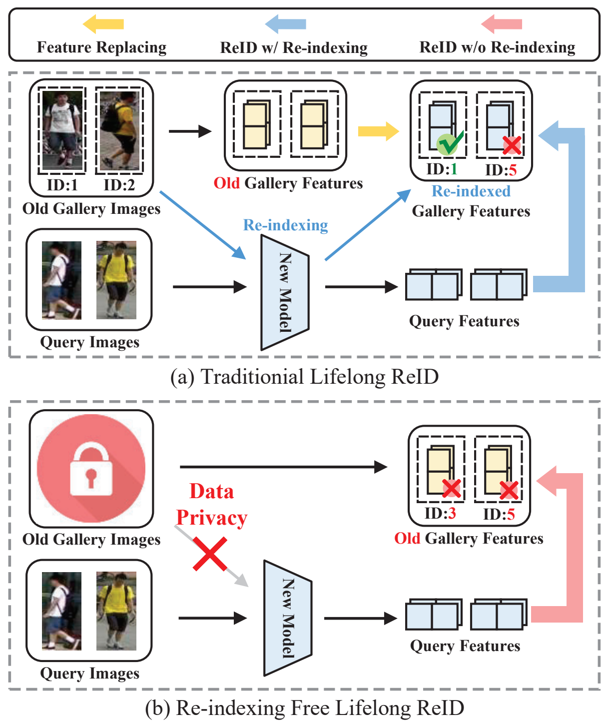
Zhenyu Cui[1], Jiahuan Zhou[1], Xun Wang[2], Manyu Zhu[2], Yuxin Peng[1]*
[1]Wangxuan Institute of Computer Technology, Peking University [2]ByteDance Inc
cuizhenyu@stu.pku.edu.cn, {jiahuanzhou,pengyuxin}@pku.edu.cn, {wangxun.2,zhumanyu}@bytedance.com

Abstract

Lifelong Person Re-identification (L-ReID) aims to learn from sequentially collected data to match a person across different scenes. Once an L-ReID model is updated using new data, all historical images in the gallery are required to be re-calculated to obtain new features for testing, known as "re-indexing". However, it is infeasible when raw images in the gallery are unavailable due to data privacy concerns, resulting in incompatible retrieval between the query and the gallery features calculated by different models, which causes significant performance degradation. In this paper, we focus on a new task called Re-indexing Free Lifelong Person Re-identification (RFL-ReID), which requires achieving effective L-ReID without re-indexing raw images in the gallery. To this end, we propose a Continual Compatible Representation (C^2R) method, which facilitates the query feature calculated by the continuously updated model to effectively retrieve the gallery feature calculated by the old model in a compatible manner. Specifically, we design a Continual Compatible Transfer (CCT) network to continuously transfer and consolidate the old gallery feature into the new feature space. Besides, a Balanced Compatible Distillation module is introduced to achieve compatibility by aligning the transferred feature space with the new feature space. Finally, a Balanced Anti-forgetting Distillation module is proposed to eliminate the accumulated forgetting of old knowledge during the continual compatible transfer. Extensive experiments on several benchmark L-ReID datasets demonstrate the effectiveness of our method against state-of-the-art methods for both RFL-ReID and L-ReID tasks. The source code of this paper is available at https://github.com/PKU-ICST-MIPL/C2R_CVPR2024.

1. Introduction

Person re-identification (ReID) aims to identify the same person across different areas at different times [42]. Existing methods [26, 32, 44] have made remarkable progress based on deep learning methods [37] and large-scale datasets [34, 41, 43]. However, their performance is often limited when training data are continuously collected from a series of different scenarios due to the well-known catastrophic forgetting challenge [3].



(a) Traditional Lifelong ReID
(b) Re-indexing Free Lifelong ReID

Figure 1. Comparison between (a) the traditional Lifelong Person Re-identification (L-ReID) task and (b) the Re-indexing Free Lifelong Person Re-identification (RFL-ReID) task.

Recently, Lifelong person ReID (L-ReID) has aroused great concerns involving acquiring knowledge from streaming data and performing well across all data [5, 18, 19, 38]. Its challenge is to balance the anti-forgetting of old knowledge with the acquisition of new knowledge. To this end, existing L-ReID methods usually adopt the exemplar replay [38] and the knowledge distillation [27] to preserve

*Corresponding author.

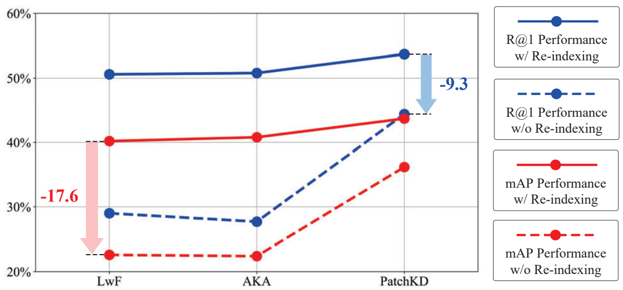


Figure 2. Performance Comparison between the traditional Lifelong Person Re-identification (L-ReID) task and the Re-indexing Free Lifelong Person Re-identification (RFL-ReID) task.

old knowledge when updating the model. The updated new model will be sequentially deployed by "re-indexing" [24] features of old data in the gallery. As shown in Fig. 1 (a), the old gallery features are replaced by the re-indexed gallery features to be compatible with the new model. However, massive incremental images in the gallery bring unbearable computational costs and prevent the re-indexing process [21, 39]. Besides, in the context of data privacy concerns [25], re-indexing is legally forbidden due to the impracticality of storing and re-indexing unauthorized raw images within privacy-sensitive scenarios [1, 4, 14], also known as Re-indexing Free Lifelong ReID (RFL-ReID), as shown in Fig. 1 (b). In this case, the domain gap between different datasets leads to the incompatibility between features calculated by the old and the new models. Consequently, as shown in Fig. 2, the performance of existing methods [11, 18, 27] often degrades significantly.

In this paper, we focus on a practical and challenging task called Re-indexing Free Lifelong Person Re-identification (RFL-ReID), which entails deploying L-ReID models continuously without re-indexing raw images in the gallery. It is extremely challenging when considering not only balancing the acquisition of old and new knowledge but also achieving the compatibility between the old and the new feature space. To this end, Compatible Training (CT) [16, 21, 23] becomes a feasible solution which promotes the compatibility. Unfortunately, most existing CT methods only focus on compatibility within the same dataset. However, the large domain gap [5, 19] between different ReID datasets typically leads to the forgetting of old knowledge, causing the feature shift problem, which seriously degrades the performance of existing CT methods.

Inspired by the above observation, we propose a Continual Compatible Representation (C^2R) scheme for RFL-ReID. The core idea of C^2R is to continuously update old features in the gallery to make it compatible with new query features. To tackle the domain shift problem, a Continual Compatible Transfer (CCT) network is proposed to update old gallery features continuously, which transfers the old feature into the new feature space by adaptively capturing the knowledge from different domains. Correspondingly, to achieve the compatibility between the transferred features and the new ones, a Balanced Compatible Distillation (BCD) module is designed to preserve the relationship between the old and the transferred features in a unified feature space. Finally, to eliminate the accumulated forgetting of old knowledge during the continuous transfer, a Balanced Anti-forgetting Distillation (BAD) module is introduced to balance the old and the new discriminative information. Overall, the main contributions of this paper can be summarized as follows:

1) This paper proposes a Continual Compatible Representation scheme to solve the challenging Re-indexing Free Lifelong Person Re-identification task, which prevents re-indexing to raw images in the old gallery.

2) A continual compatible transfer network is designed to adaptively capture the old and the new knowledge, which promotes the updating of old gallery features.

3) To achieve compatibility of the old and the new data, a balanced compatible distillation module is designed to balance the relationship between the old and the new features in a unified feature space.

4) A balanced anti-forgetting distillation module is introduced to balance the acquisition of the old and the new knowledge without rehearsing the old data.

2. Related Work

2.1. Lifelong Person Re-identification

Lifelong Person Re-identification (L-ReID) aims to use streaming data from different domains to train a ReID model to continuously match the same person across all domains [18]. It requires continuously accumulating knowledge from new data and preventing catastrophic forgetting of knowledge learned in old data. To this end, rehearsal-based methods rehearsed old data for the anti-forgetting of old knowledge [5, 35, 38]. Yu et al. [38] proposed a knowledge refreshing and consolidation framework to achieve both positive forward and backward transfers, which simultaneously promotes person matching in both old and new domains. However, limited by data privacy concerns, the old data is usually difficult to rehearse for L-ReID. Therefore, some methods focus on developing rehearsal-free methods [18, 19, 27, 33]. Sun et al. [27] proposed a patch-based knowledge distillation method, which can mitigate catastrophic forgetting problems by preserving patch-level knowledge within individual patch features and mutual relations. Despite achieving some progress, existing methods require re-indexing old data in the gallery after training on new data to adapt to the updated model. In this paper, we focus on a new and challenging task called RFL-ReID, which prohibits re-indexing the old data in the gallery due to data privacy issues.

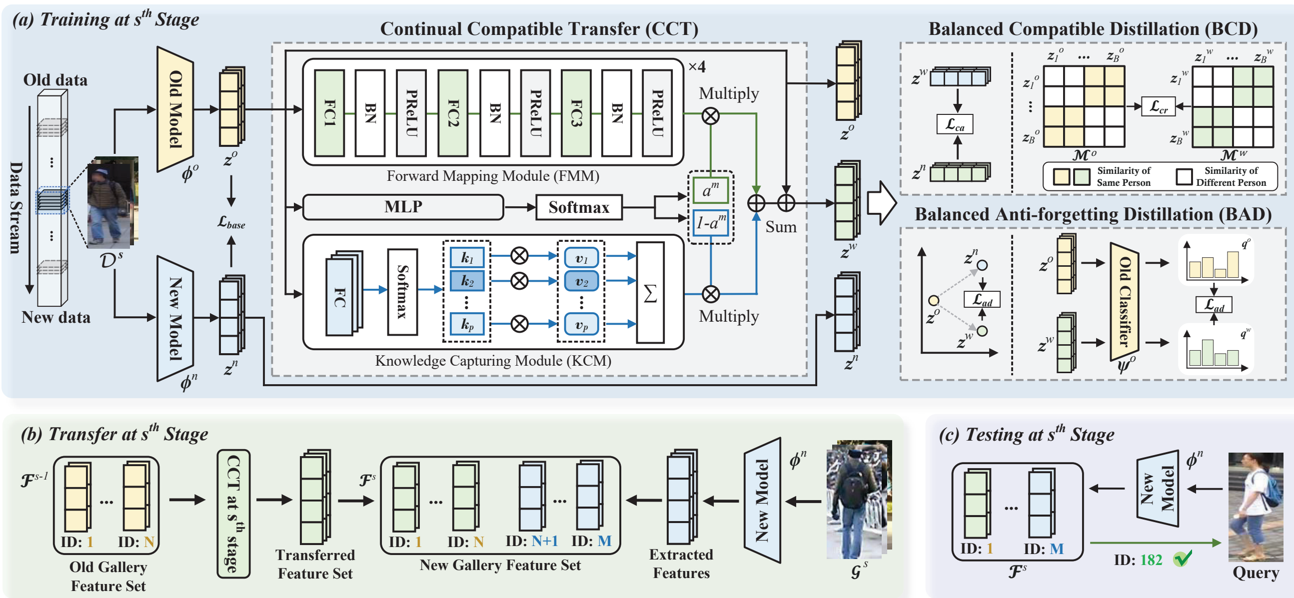


Figure 3. The architecture of our proposed Continual Compatible Representation (C²R) method. Our C²R consists of a Continual Compatible Transfer (CCT) network, a Balanced Compatible Distillation (BCD) module, and a Balanced Anti-forgetting Distillation (BAD) module. In the training phase, all the above components will be trained, and the CCT network will be used to update the old feature set, thereby achieving re-indexing free lifelong person re-identification.

2.2. Compatible Training

Compatible Training (CT) aims to achieve the compatibility between the old model and the new model without re-indexing, which encourages the feature calculated by the new model to be closer to the same object feature calculated by the old model. Existing CT methods can be categorized into two branches: backward CT [16, 24, 39] and forward CT methods [21, 30, 45]. Pan et al. [16] proposed an adversarial learning-based backward CT method, which optimized the distribution discrepancy and improved the backward compatibility of the updated model. Ramanujan et al. [21] proposed a flexible forward CT method by employing side-information to facilitate the updating of the new model. However, most existing CT methods only achieve compatibility between the old and the new models updated using the same dataset, ignoring the requirement to be compatible with multiple different datasets continuously. To this end, Wan et al. [30] proposed a long-term visual search framework to realize CT in the continual learning scenario, which allows new classes to occur in the new data. However, it relies heavily on old data and old features to achieve the compatibility, which is infeasible for the privacy-sensitive scenario.

Different from these methods, we propose a continual compatible representation scheme called C²R, which achieves leading L-ReID performance when old data in the gallery cannot be re-indexed in the practical privacy-sensitive scenario.

3. Method

3.1. Problem Formulation

In this paper, we focus on a practical and challenging task called Re-indexing Free Lifelong Person Re-identification (RFL-ReID). Formally, given the sequentially collected datasets $\mathcal{D} = \{\mathcal{D}^1, \mathcal{D}^2, ..., \mathcal{D}^S\}$, where S denotes the total number. The s^{th} dataset $\mathcal{D}^s = \{\mathcal{T}^s, \mathcal{G}^s\}$ contains a training set $\mathcal{T}^s$ and the corresponding gallery set $\mathcal{G}^s$ without overlapping person identities. Each $\mathcal{T}^s = \{x_i^s, y_i^s\}_{i=1}^{N^s}$ contains N^s training images x_i^s and the corresponding identity labels y_i^s. Our goal is to train a feature extraction model $\phi^S(\cdot) : x_i \rightarrow \boldsymbol{z}_i$ to extract the feature $\boldsymbol{z}_i$ for each input image x_i. $\phi^S(\cdot)$ is expected to maximize the similarity between the images belonging to the same person, thereby performing well on the overall retrieval performance of all S tasks. Particularly, when training on the s^{th} dataset $\mathcal{D}^s$, the previous $s-1$ datasets $\{\mathcal{D}^1, ..., \mathcal{D}^{s-1}\}$ are absolutely unavailable due to data privacy concerns, including the training sets $\{\mathcal{T}^j\}_{j=1}^{s-1}$ and the gallery sets $\{\mathcal{G}^j\}_{j=1}^{s-1}$. Therefore, each image in $\mathcal{G}^j$ can only be extracted once based on the model $\phi^j(\cdot)$ within the j^{th} training stage.

3.2. Overview

As shown in Fig. 3, our proposed Continual Compatible Representation (C²R) method consists of a Continual Compatible Transfer (CCT) network, a Balanced Compatible Distillation (BCD) module, and a Balanced Anti-forgetting

Distillation (BAD) module. At the beginning of the s^{th} training stage ($s \geq 2$), a copy of the model ϕ^o and the classifier ψ^o are frozen to serve as the old knowledge. Then, we train the model with the CCT network based on the BCD and the BAD modules, as shown in Fig. 3 (a). After the training of the s^{th} stage, we reform a new gallery feature set $\mathcal{F}^s$ by updating the transferred features in the old gallery feature set $\mathcal{F}^{s-1}$ using our proposed CCT network, and extracting the features of the current gallery set $\mathcal{G}^s$ using ϕ^s (Fig. 3 (b)). When testing at the s^{th} stage, we rank each feature in $\mathcal{F}^s$ with the query feature calculated by ϕ^s (Fig. 3 (c)) to validate the ReID performance on all s datasets.

3.3. Lifelong Person Re-identification Baseline

In this section, we start by presenting a lifelong person re-identification baseline. The baseline model consists of a feature extraction model $\phi(\cdot)$ and a classifier head $\psi(\cdot)$, where $\phi(\cdot)$ extracts a feature z_i for each input image x_i, and $\psi(\cdot)$ predicts the probability p_i of each person's identity.

Considering that the RFL-ReID task aims to continuously match the same person across the sequentially given datasets, we employ a cross-entropy loss $\mathcal{L}_{ce}$ [13] and a triplet loss $\mathcal{L}_{trip}$ [8] to learn a discriminative representation for each person in the current dataset $\mathcal{D}^s$. However, despite certain performance improvements in $\mathcal{D}^s$, it has been verified to cause the catastrophic forgetting problem on the previous $(s-1)$ datasets [3]. To balance the acquisition of new knowledge and the anti-forgetting of old knowledge, we introduce an anti-forgetting loss $\mathcal{L}_{pkd}$ [27] based on image patch knowledge distillation to build a basic L-ReID model. Therefore, the overall loss function for our baseline model is formulated as:

$$\mathcal{L}_{base} = \mathcal{L}_{ce} + \mathcal{L}_{trip} + \mathcal{L}_{pkd}. \tag{1}$$

The above baseline method achieves a performance trade-off between new and old datasets by re-indexing data in old galleries using the new model. However, when the old data is unavailable to the new model due to data privacy concerns, the performance will be significantly degraded due to the incompatibility of the features output by the old model and the new model. To tackle the above issue, we will detail present each component of our C^{2}R scheme in the following sections.

3.4. Continual Compatible Transfer Network

To tackle the domain shift problem, we design a Continual Compatible Transfer (CCT) network to transfer the old feature into the new feature space continuously. To this end, our CCT is designed as a two-stream network to capture transfer-related knowledge from different domains.

Specifically, we first design a Knowledge Capturing Module (KCM) module to capture knowledge from new domains. Let $\boldsymbol{z}^o \in \mathbb{R}^C$ be the feature output by the old model $\phi^o(\cdot)$ based on the sample x^s, $\boldsymbol{v}_p \in \mathbb{R}^{P \times C}$ be a learnable new knowledge prototype set with length P. In order to enable $\boldsymbol{z}^o$ to capture new knowledge for compatibility, we first calculate a capturing probability $\boldsymbol{k}^o \in \mathbb{R}^P$ for each new knowledge prototype:

$$\boldsymbol{k}^o = \delta(g_c(\widetilde{\boldsymbol{z}}^o)), \tag{2}$$

where $g_c(\cdot)$ denotes a three-layer fully connected network, $\delta(\cdot)$ denotes the softmax function, and $\widetilde{(\cdot)}$ denotes the l2-normalized function. Then, we combine the captured new knowledge with a weighted summation of the prototypes:

$$\boldsymbol{z}^t = \sum_{p=1}^{P} \boldsymbol{k}_i^o \cdot \boldsymbol{v}_p. \tag{3}$$

Sequentially, a parallel Forward Mapping Module (FMM) is designed to directly map the old feature to the new one, thereby preserving the old knowledge from old domains. The designed mapping network consists of sequential fully connected layers, Batch Normalization [9] layers, and PReLU [6] layers. The first fully connected layer reduces the channel number to C_0, while the last layer recovers it back to the original number, which keeps the representation capability while saving redundant operations:

$$\boldsymbol{z}^m = g_s(\widetilde{\boldsymbol{z}}^o). \tag{4}$$

Finally, CCT takes the input $\widetilde{\boldsymbol{z}}^o$ to generate a factor a^m to adaptively balance the above two features ($\boldsymbol{z}^t$ and $\boldsymbol{z}^m$) obtained by the KCM module and FMM module. Meanwhile, the original features are added by the residual skip-connection to alleviate the gradient vanishing issue [31]. The transferred result $\boldsymbol{z}^w$ of CCT network can be formulated as follows:

$$\boldsymbol{z}^w = (1 - a^m) \cdot \boldsymbol{z}^t + a^m \cdot \boldsymbol{z}^m + \widetilde{\boldsymbol{z}}^o. \tag{5}$$

Following [14], we leverage 4 cascaded CCT networks to build up a strong transfer network, which sequentially transfers features from previous stages to the current stage.

3.5. Balanced Compatible Distillation

To make the transferred features compatible with the new one, we design a Balanced Compatible Distillation (BCD) module to simultaneously preserve the relationships of the old and the transferred features in the new feature space.

Given a mini-batch of training samples $\{x_i^s, y_i^s\}_{i=1}^B$, where B is the batch size, and a new feature $\boldsymbol{z}_i^n$ calculated by the new model $\phi(\cdot)^n$. We start by directly aligning the above features with l2-loss:

$$\mathcal{L}_{ca} = -\frac{1}{B} \sum_{i=1}^{B} ||\widetilde{\boldsymbol{z}}_i{}^n - \widetilde{\boldsymbol{z}}_i{}^w||_2. \tag{6}$$

The above alignment loss promotes compatibility of new features in the new feature space, while ignoring the relationship between the old features, thus limiting the ReID performance on old data. Therefore, we propose a relationship distillation loss to overcome the above issue.

Firstly, we construct an old similarity matrix $\mathcal{M}^o \in \mathbb{R}^{B\times B}$, which is calculated by the affinity between each two features in $\{\widetilde{\boldsymbol{z}}_i^o\}_{i=1}^B$ to represent the relationship between the corresponding images in the old feature space:

$$\mathcal{M}_{i,j}^o = \frac{e^{\langle \widetilde{\boldsymbol{z}}_i^o, \widetilde{\boldsymbol{z}}_j^o \rangle}}{\sum_{k=1}^B e^{\langle \widetilde{\boldsymbol{z}}_i^o, \widetilde{\boldsymbol{z}}_k^o \rangle}}, (i, j \in [1, B]), \tag{7}$$

where $\langle\cdot,\cdot\rangle$ denotes the cosine similarity. Similarly, a transferred similarity matrix $\mathcal{M}^w$ is also constructed to represent the relationship between the transferred features $\{\widetilde{\boldsymbol{z}}_i^w\}_{i=1}^B$. Apparently, the alignment in Eq. (6) will promote higher similarity between samples with the same identity in $\mathcal{M}^w$, which will conflict with the corresponding similarity in $\mathcal{M}^o$. Therefore, we remove the similarities between samples belonging to the same identity to avoid the above conflicts for relationship distillation:

$$\mathcal{M}_{i,j}^x = \begin{cases} \mathcal{M}_{i,j}^x, ID(i) \neq ID(j) \\ 0, \qquad ID(i) = ID(j), \end{cases} \tag{8}$$

where $ID(\cdot)$ denotes the identity of the given sample, and $x \in o, w$. Next, a widely-used knowledge distillation loss $\mathcal{L}_{cr}$ based on Kullback-Leibler (KL) divergence is imposed to distil the relationship between old features as follows:

$$\mathcal{L}_{cr} = -\frac{1}{B}\sum_{i=1}^{B} \sigma(M_{i,:}^o) log(\frac{\sigma(M_{i,:}^o)}{\sigma(M_{i,:}^w)}), \tag{9}$$

where $\sigma(\cdot)$ denotes the row-wise l1-normalized function.

In all, the loss function for the BCD module can be calculated as follows:

$$\mathcal{L}_{bcd} = \mu_1\mathcal{L}_{ca} + \mu_2\mathcal{L}_{cr}, \tag{10}$$

where μ_1 and μ_2 denote two hyper-parameters to balance our proposed balanced compatible distillation.

3.6. Balanced Anti-forgetting Distillation

Although the proposed BCD module achieves the compatibility by selectively distilling the relationship, it ignores the accumulated forgetting of old knowledge during the continual compatible transfer, thereby reducing the inter-class discrimination of old data. Therefore, we propose a Balanced Anti-forgetting Distillation (BAD) module to preserve the discriminative information within the old and new features in the new feature space.

To this end, we employ the old classifier head $\psi^o(\cdot)$ to extract and distil the discriminative information. Therefore, we extract old identity logits $\boldsymbol{q}^o$ of the old feature $\boldsymbol{z}^o$ as the theoretical distribution of the discriminative information:

$$\boldsymbol{q}^o = \delta(\psi^o(\boldsymbol{z}^o)). \tag{11}$$

Considering that the transferred feature $\boldsymbol{z}^w$ and $\boldsymbol{z}^o$ come from different datasets, it is challenging to extract consistent discriminative information directly using $\psi^o(\cdot)$. To tackle the above problem, we perform an inverse l2-normalization on $\boldsymbol{z}^w$ so that it can be extracted as the real distribution of the transferred discriminative information:

$$\boldsymbol{q}^w = \delta(\psi^o(\widetilde{\boldsymbol{z}}^w \cdot ||\boldsymbol{z}^o||_2)). \tag{12}$$

Sequentially, we use an anti-forgetting loss $\mathcal{L}_{ad}$ to preserve the old discriminative knowledge, which can be calculated as follows:

$$\mathcal{L}_{ad} = -\frac{1}{B}\sum_{i=1}^{B} \boldsymbol{q}_i^o log(\frac{\boldsymbol{q}_i^o}{\boldsymbol{q}_i^w}). \tag{13}$$

Despite achieving the anti-forgetting, the distillation in Eq. (13) will also limit the discrimination of the transferred feature in the new feature space. Therefore, we introduce a discriminative consistency-based loss $\mathcal{L}_{dc}$ to balance the discrimination for both transferred and new features:

$$\mathcal{L}_{dc} = -\frac{1}{B}\sum_{i=1}^{B}(1 - \cos\langle \frac{\widetilde{\boldsymbol{z}}_i^{\,w} - \widetilde{\boldsymbol{z}}_i^{\,o}}{||\widetilde{\boldsymbol{z}}_i^{\,w} - \widetilde{\boldsymbol{z}}_i^{\,o}||_2}, \frac{\widetilde{\boldsymbol{z}}_i^{\,n} - \widetilde{\boldsymbol{z}}_i^{\,o}}{||\widetilde{\boldsymbol{z}}_i^{\,n} - \widetilde{\boldsymbol{z}}_i^{\,o}||_2}\rangle). \tag{14}$$

The overall loss function for the BAD module can be calculated as follows:

$$\mathcal{L}_{bad} = \mu_3\mathcal{L}_{ad} + \mu_4\mathcal{L}_{dc}, \tag{15}$$

where μ_3 and μ_4 denote two hyper-parameters when training the model.

3.7. Objective Function

Finally, the total loss $\mathcal{L}$ of C^2R is calculated as follows:

$$\mathcal{L} = \mathcal{L}_{base} + \mathcal{L}_{bcd} + \mathcal{L}_{bad}. \tag{16}$$

Through the joint optimizing of $\mathcal{L}$, our C^2R achieves the compatibility between the old and the new models and prevents re-indexing to raw images in the old gallery.

4. Experiments

4.1. Datasets and Evaluation Metrics

Datasets. To verify the effectiveness of our method, we conduct extensive experiments on five benchmark lifelong person ReID datasets, including Market-1501 [41], CUHK-SYSU [36], DukeMTMC-ReID [43], MSMT17-V2 [34] and CUHK03 [12]. For CUHK-SYSU, we follow the

Task	Method	Market-1501		CUHK-SYSU		DukeMTMC-ReID		MSMT17-V2		CUHK03		Average	
		mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
	JointTrain	68.1	85.2	81.4	83.8	60.4	75.7	24.6	48.9	42.7	43.6	55.4	67.5
L-ReID	SPD [28]	35.6	61.2	61.7	64.0	27.5	47.1	5.2	15.5	**42.2**	**44.3**	34.4	46.4
	LwF [11]	56.3	77.1	72.9	75.1	29.6	46.5	6.0	16.6	36.1	37.5	40.2	50.6
	CRL [40]	58.0	78.2	72.5	75.1	28.3	45.2	6.0	15.8	37.4	39.8	40.5	50.8
	AKA [18]	58.1	77.4	72.5	74.8	28.7	45.2	6.1	16.2	38.7	40.4	40.8	50.8
	MEGE [20]	39.0	61.6	73.3	76.6	16.9	30.3	4.6	13.4	36.4	37.1	34.0	43.8
	PatchKD [27]	68.5	85.7	75.6	78.6	**33.8**	**50.4**	6.5	17.0	34.1	36.8	43.7	**53.7**
	Ours	**69.0**	**86.8**	**76.7**	**79.5**	33.2	48.6	**6.6**	**17.4**	35.6	36.2	**44.2**	**53.7**
RFL-ReID	LwF* [11]	39.1	58.0	40.0	40.7	7.8	15.3	2.6	7.1	23.3	23.9	22.6	29.0
	AKA* [18]	36.1	52.2	38.6	37.6	7.6	13.8	3.1	8.3	26.5	26.5	22.4	27.7
	CVS* [30]	38.8	55.6	49.0	49.7	19.3	30.0	4.6	11.5	24.7	24.7	27.3	34.3
	PatchKD* [27]	61.4	78.4	57.8	59.0	20.8	34.4	5.1	12.8	36.0	37.6	36.2	44.4
	Ours	**62.7**	**79.7**	**64.4**	**66.3**	**26.7**	**41.7**	**6.8**	**15.7**	**37.2**	**37.6**	**39.5**	**48.2**

Table 1. Performance on training *Order-1*: Market-1501→CUHK-SYSU→DukeMTMC-ReID→MSMT17-V2→CUHK03.

Task	Method	DukeMTMC-ReID		MSMT17-V2		Market-1501		CUHK-SYSU		CUHK03		Average	
		mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
	JointTrain	60.4	75.7	24.6	48.9	68.1	85.2	81.4	83.8	42.7	43.6	55.4	67.5
L-ReID	SPD [28]	28.5	48.5	3.7	11.5	32.3	57.4	62.1	65.0	**43.0**	**45.2**	33.9	45.5
	LwF [11]	42.7	61.7	5.1	14.3	34.4	58.6	69.9	73.0	34.1	34.1	37.2	48.4
	CRL [40]	43.5	63.1	4.8	13.7	35.0	59.8	70.0	72.8	34.5	36.8	37.6	49.2
	AKA [18]	42.2	60.1	5.4	15.1	37.2	59.8	71.2	73.9	36.9	37.9	38.6	49.4
	MEGE [20]	21.6	35.5	3.0	9.3	25.0	49.8	69.9	73.1	34.7	35.1	30.8	40.6
	PatchKD [27]	58.3	74.1	6.4	17.4	**43.2**	**67.4**	74.5	76.9	33.7	34.8	43.2	54.1
	Ours	**59.7**	**75.0**	**7.3**	**19.2**	42.4	66.5	**76.0**	**77.8**	37.8	39.3	**44.7**	**55.6**
RFL-ReID	LwF* [11]	15.0	22.9	1.2	3.2	9.5	19.4	38.8	37.5	20.2	19.6	16.9	20.5
	AKA* [18]	11.1	15.1	1.3	3.2	13.4	27.3	35.9	34.7	25.2	25.6	17.4	21.2
	CVS* [30]	29.0	41.9	3.5	9.4	30.7	49.6	60.0	61.2	28.5	29.9	30.3	38.4
	PatchKD* [27]	46.5	60.9	4.0	10.4	31.1	50.5	63.0	64.0	35.8	36.6	36.1	44.5
	Ours	**48.4**	**63.6**	**6.2**	**14.9**	**37.0**	**55.6**	**67.4**	**68.4**	**39.2**	**39.5**	**39.7**	**48.4**

Table 2. Performance on training *Order-2*: DukeMTMC-ReID→MSMT17-V2→Market-1501→CUHK-SYSU→CUHK03.

same pre-processing and evaluation of GwFReID [35], which crops the person images with the ground-truth person bounding box annotation and re-organized as a ReID dataset with corresponding identities. To simulate the lifelong person ReID process in the real scenarios, we evaluate our method by employing two training orders used in [18], including *Order-1*: Market-1501 → CUHK-SYSU → DukeMTMC-ReID → MSMT17-V2 → CUHK03 and *Order-2*: DukeMTMC-ReID → MSMT17-V2 → Market-1501 → CUHK-SYSU → CUHK03.

Evaluation Metrics. We employ mean Average Precision (mAP) [41] and Rank@1 accuracy (R@1) [15] to evaluate our method on each datasets. The above metrics of our method are reported after sequentially training all datasets. Besides, the corresponding mean accuracy is also reported to evaluate the overall performance on all tasks. Besides, the Average Forgetting (AF) [2] on the above metrics are reported to evaluate the anti-forgetting performance of our method, which averages the difference between the highest accuracy and the final accuracy on each dataset throughout the lifelong learning process.

4.2. Implementation Details

Our proposed C^{2}R is implemented with PyTorch [17] on NVIDIA A40 GPUs. We adopt a ResNet-50 [7] network pre-trained on ImageNet [22] as our backbone network. In each training stage, we train the model for 50 epochs with 150 iterations. We use Adam [10] optimizer for training. The learning rate is set to 3.5×10^{-4} initially, which decayed by 0.1 at the 25^{th} and the 35^{th} epochs. We set the hyper-parameters μ_1, μ_2, μ_3, and μ_4 as 50, 1, 0.01 and 0.05, respectively. The prototype length P and the channel number C_0 are empirically set to 16 and 32. Input images are resized to 256×128. The batch size is set to 128, with 4 images for each identity. Following [27], each dataset is randomly sampled with 500 identities to alleviate the problem of unbalanced class numbers among different datasets for fairness comparison.

Method	LwF	AKA	CVS	PatchKD	**Ours**
AF(mAP)	24.8	26.4	21.0	16.5	**13.9**
AF(R@1)	30.8	35.0	25.7	18.5	**14.5**

Table 3. AF performance of C^2R compared with existing methods. (The lower the AF is, the less the model forgets.)

4.3. Comparison with State-of-the-arts Methods

In this section, we compared our C^2R with several state-of-the-art methods on both RFL-ReID and traditional L-ReID tasks, including SPD [28], LwF [11], CRL [40], AKA [18], MEGE [20], PatchKD [27], and CVS [30], where CSV is a continual compatible training method and the others are L-ReID methods. The best results are **bolded**, and the second-best results are underlined.

Comparison on the RFL-ReID task. Tab. 1 and Tab. 2 summarize the results of our method compared to existing methods, which are reproduced by the released code and denoted as $*$. Compared with these methods, our C^2R achieves the highest average mAP and R@1 accuracy by 39.5%/48.2% on *Order-1* and 39.7%/48.4% on *Order-2*, leading PatchKD by 3.3%/3.8% and 3.6%/3.9%, respectively. This is because our C^2R method can achieve compatible transfer for old gallery features by balancing the anti-forgetting of old knowledge with the compatibility to the new model, thereby updating old features without re-indexing old data effectively.

In addition, Tab. 3 reports the AF performance of mAP and R@1 on *Order-1*. It can be seen that our C^2R achieves the lowest AF on both mAP and R@1 accuracy, which is lower than PatchKD by 2.6% and 4.0%, respectively. Notably, compared to CVS, our C2R does not introduce additional old data while achieving the best RFL-ReID accuracy and the best AF performance. Specifically, as shown in Fig. 4, our proposed C^2R achieves the best average performance on the seen datasets in each stage. The above results intuitively demonstrate that our proposed C^2R outperforms SOTA methods on average R@1 and mAP at each stage.

The above results imply that our C^2R can adaptively map old features and capture new knowledge, thus balancing compatibility with new feature spaces by preserving old discriminative knowledge.

Comparison on the traditional L-ReID task. We further evaluate our method on the traditional L-ReID task. As shown in Tab. 1 and Tab. 2, our proposed C^2R also achieves state-of-the-art performance on both two training orders, which leads PatchKD on mAP accuracy by 0.5% and 1.5% on *Order-1* and *Order-2*, respectively. The above results imply that our method achieves the best RFL-ReID performance without compromising the performance of traditional L-ReID. Therefore, our C^2R has better generalization in both general and privacy-sensitive scenarios.

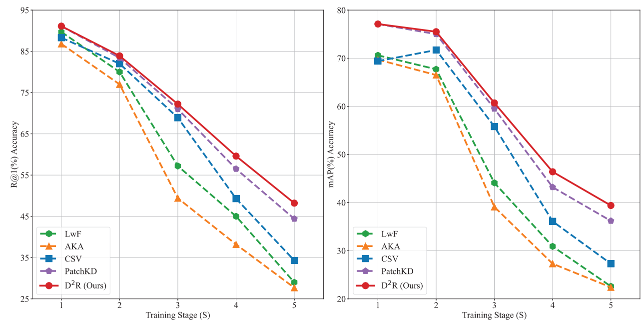


Figure 4. Lifelong learning performance comparison in RFL-ReID task on mAP and R@1 accuracy.

a^m	0.0	0.2	0.4	0.6	0.8	1.0	**Ours**
mAP	38.2	38.7	39.0	38.4	38.9	38.7	**39.5**
R@1	46.3	46.9	47.6	47.5	47.8	47.7	**48.2**

Table 4. Ablation study on different balancing factors in continual compatible transfer network.

4.4. Ablation Studies

In this section, we conduct ablation studies on *Order-1* to evaluate the effectiveness of each component in our C^2R. Our Base method is the baseline introduced in Sec. 3.3.

Effectiveness of CCT network. We compared the adaptively balancing strategy in our continual compatible transfer network with a fixed balancing strategy, which introduces fixed values for a^m in Eq. (5). As shown in Tab. 4, our method achieves the best mAP and R@1 accuracy among the fixed balancing strategies. It indicates that by adaptively generating the most appropriate fusion factor through different old features, our CCT network balances the forward mapping of old knowledge and the knowledge capturing of new knowledge, verifying its superior effectiveness.

Effectiveness of BCD module. As shown in Tab. 5, we evaluate the two important loss functions $\mathcal{L}_{ca}$ (Eq. (6)) and $\mathcal{L}_{cr}$ (Eq. (9)) in the balanced compatible distillation module. It can be seen that compared with the Base method, $\mathcal{L}_{ca}$ improves the average mAP/R@1 by 1.7% and 1.2%, and the full version of the BCD module improves by 1.9% and 2.3%. The above results indicate that the alignment in $\mathcal{L}_{ca}$ and the relationship distillation in $\mathcal{L}_{cr}$ are both important for RFL-ReID. The former ensures that old features can continue to be compatible with the new ones, while the latter preserves the relationship within the old feature space. Finally, the BCD module achieves promising improvement by modelling the above functions in a balanced manner.

Effectiveness of BAD module. We further evaluate two main components $\mathcal{L}_{ad}$ (Eq. (13)) and $\mathcal{L}_{dc}$ (Eq. (14)) in the balanced anti-forgetting module. As shown in Tab. 5, $\mathcal{L}_{ad}$ brings an improvement by 1.1% on average mAP accuracy. When combined with $\mathcal{L}_{dc}$, the accuracy further improved by 0.3%. It indicates that by distilling the discriminative

Base	CCT	BCD		BAD		mAP	R@1
		$\mathcal{L}_{ca}$	$\mathcal{L}_{cr}$	$\mathcal{L}_{ad}$	$\mathcal{L}_{dc}$		
✓	-	-	-	-	-	36.2	44.4
✓	✓	✓	-	-	-	37.9	45.6
✓	✓	✓	✓	-	-	38.1	46.7
✓	✓	✓	✓	✓	-	39.2	47.1
✓	✓	✓	✓	✓	✓	**39.5**	**48.2**

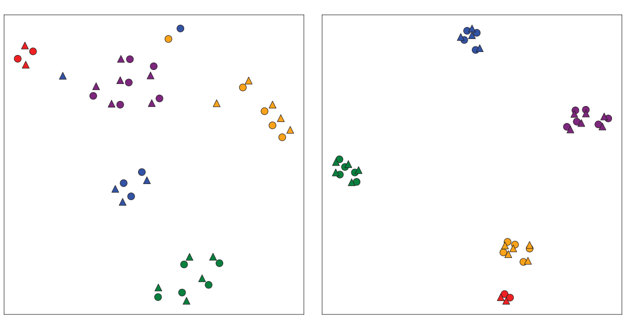
Table 5. Ablation studies of each component in C^{2}R on *Order-1*.

information within the old feature, BAD balances the anti-forgetting of the old discrimination knowledge with the acquisition of the new knowledge and thus plays an important role in our C^{2}R method.

Hyper-parameter Study. There are 4 hyper-parameters for the training of our C^{2}R method, *i.e.* μ_1, μ_2, μ_3, and μ_4. Therefore, we evaluate the effect of the above parameters respectively, as shown in Fig. 6. It can be seen that a slight or excessive μ_1 will eliminate the alignment of the old feature to the new feature. In addition, as μ_2 increases, the relationship between old features will prevent the model from learning new knowledge. Therefore, the moderate μ_1(=50) and μ_2(=1) are chosen in the BCD module to balance the relationship between the old and new features. Similarly, we set a moderate μ_3(=0.01) and μ_4(=0.05) to balance the discriminative information in the old and new features.

4.5. Visualization

To intuitively verify the effectiveness of our C^{2}R, we use t-SNE [29] to visualize and compare the gallery feature and query feature calculated by our method with the Base method. The features are randomly selected from the five benchmark datasets. As shown in Fig. 5, the features calculated by our C^{2}R achieve tighter intra-class representation while retaining more inter-class discriminability. The above results indicate that our C^{2}R can preserve more discriminative information after transferring old gallery features, which can be further aligned by new query features to achieve RFL-ReID.



(a) Base (b) our C^{2}R

Figure 5. TSNE results of our C^{2}R compared with Base method. Different colours represent different identities, while the circles and the triangles represent gallery and query features, respectively.

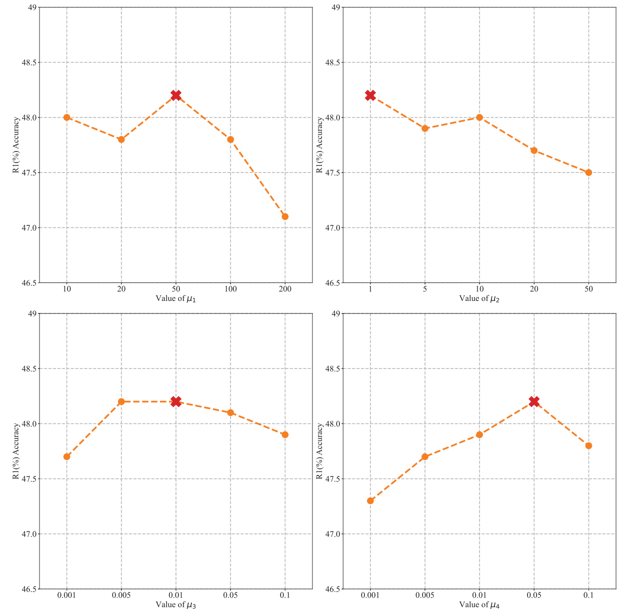


Figure 6. The influence of hyper-parameters in our C^{2}R.

4.6. Limitation and Future Work

Our C^{2}R can support L-ReID without re-indexing old data in the gallery. However, it is difficult to capture enough new knowledge to facilitate feature transfer and updating when the amount of new data is limited. Therefore, it is worthy of future research on efficient transfer under a few-shot scenario. In addition, when there is noise in new data, how to utilize old knowledge to eliminate such noise to achieve selective transfer should be studied in the future.

5. Conclusion

In this paper, we focus on a practical and challenging task called Re-indexing Free Lifelong Person Re-identification (RFL-ReID), which prohibits the re-indexing of raw images in the gallery due to data privacy concerns. To this end, we propose a Continual Compatible Representation (C^{2}R) method, which introduces a Continual Compatible Transfer (CCT) network to continuously transfer old gallery features to the new feature space. Besides, a Balanced Compatible Distillation module and a Balanced Anti-forgetting Distillation module are proposed to balance the anti-forgetting of old knowledge with the compatibility to the new model. Extensive experiments on five widely-used benchmark L-ReID datasets verify the effectiveness of our method on the RFL-ReID task while maintaining the state-of-the-art performance on the general L-ReID scenario.

Ackowledgements

This work was supported by the National Natural Science Foundation of China (61925201, 62132001, 62376011).

References

[1] Shafiq Ahmad, Gianluca Scarpellini, Pietro Morerio, and Alessio Del Bue. Event-driven re-id: A new benchmark and method towards privacy-preserving person re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 459–468, 2022. 2

[2] Arslan Chaudhry, Puneet K Dokania, Thalaiyasingam Ajanthan, and Philip HS Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *Proceedings of the European conference on computer vision (ECCV)*, pages 532–547, 2018. 6

[3] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3366–3385, 2021. 1, 4

[4] Zekeriya Erkin, Martin Franz, Jorge Guajardo, Stefan Katzenbeisser, Inald Lagendijk, and Tomas Toft. Privacy-preserving face recognition. In *Privacy Enhancing Technologies: 9th International Symposium, PETS 2009, Seattle, WA, USA, August 5-7, 2009. Proceedings 9*, pages 235–253. Springer, 2009. 2

[5] Wenhang Ge, Junlong Du, Ancong Wu, Yuqiao Xian, Ke Yan, Feiyue Huang, and Wei-Shi Zheng. Lifelong person re-identification by pseudo task knowledge preservation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 688–696, 2022. 1, 2

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1026–1034, 2015. 4

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[8] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 4

[9] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015. 4

[10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[11] Dangwei Li, Xiaotang Chen, Zhang Zhang, and Kaiqi Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 384–393, 2017. 2, 6, 7

[12] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 152–159, 2014. 5

[13] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019. 4

[14] Qiang Meng, Chixiang Zhang, Xiaoqiang Xu, and Feng Zhou. Learning compatible embeddings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9939–9948, 2021. 2, 4

[15] Hyeonjoon Moon and P Jonathon Phillips. Computational and performance aspects of pca-based face-recognition algorithms. *Perception*, 30(3):303–321, 2001. 6

[16] Tan Pan, Furong Xu, Xudong Yang, Sifeng He, Chen Jiang, Qingpei Guo, Feng Qian, Xiaobo Zhang, Yuan Cheng, Lei Yang, et al. Boundary-aware backward-compatible representation via adversarial learning in image retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15201–15210, 2023. 2, 3

[17] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 6

[18] Nan Pu, Wei Chen, Yu Liu, Erwin M Bakker, and Michael S Lew. Lifelong person re-identification via adaptive knowledge accumulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7901–7910, 2021. 1, 2, 6, 7

[19] Nan Pu, Yu Liu, Wei Chen, Erwin M Bakker, and Michael S Lew. Meta reconciliation normalization for lifelong person re-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 541–549, 2022. 1, 2

[20] Nan Pu, Zhun Zhong, Nicu Sebe, and Michael S Lew. A memorizing and generalizing framework for lifelong person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 6, 7

[21] Vivek Ramanujan, Pavan Kumar Anasosalu Vasu, Ali Farhadi, Oncel Tuzel, and Hadi Pouransari. Forward compatible training for large-scale embedding retrieval systems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19386–19395, 2022. 2, 3

[22] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015. 6

[23] Seonguk Seo, Mustafa Uzunbas, Bohyung Han, Xuefei Cao, Joena Zhang, Taipeng Tian, and Ser-Nam Lim. Metric compatible training for online backfilling in large-scale retrieval. In *ICML Workshop on Localized Learning (LLW)*, 2023. 2

[24] Yantao Shen, Yuanjun Xiong, Wei Xia, and Stefano Soatto. Towards backward-compatible representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6368–6377, 2020. 2, 3

[25] Shupeng Su, Binjie Zhang, Yixiao Ge, Xuyuan Xu, Yexin Wang, Chun Yuan, and Ying Shan. Privacy-preserving model upgrades with bidirectional compatible training in image retrieval. *arXiv preprint arXiv:2204.13919*, 2022. 2

[26] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 480–496, 2018. 1

[27] Zhicheng Sun and Yadong Mu. Patch-based knowledge distillation for lifelong person re-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 696–707, 2022. 1, 2, 4, 6, 7

[28] Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1365–1374, 2019. 6, 7

[29] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9 (11), 2008. 8

[30] Timmy ST Wan, Jun-Cheng Chen, Tzer-Yi Wu, and Chu-Song Chen. Continual learning for visual search with backward consistent feature embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16702–16711, 2022. 3, 6, 7

[31] Chien-Yi Wang, Ya-Liang Chang, Shang-Ta Yang, Dong Chen, and Shang-Hong Lai. Unified representation learning for cross model compatibility. *arXiv preprint arXiv:2008.04821*, 2020. 4

[32] Faqiang Wang, Wangmeng Zuo, Liang Lin, David Zhang, and Lei Zhang. Joint learning of single-image and cross-image representations for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1288–1296, 2016. 1

[33] Kai Wang, Chenshen Wu, Andy Bagdanov, Xialei Liu, Shiqi Yang, Shangling Jui, and Joost van de Weijer. Positive pair distillation considered harmful: Continual meta metric learning for lifelong object re-identification. *arXiv preprint arXiv:2210.01600*, 2022. 2

[34] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 79–88, 2018. 1, 5

[35] Guile Wu and Shaogang Gong. Generalising without forgetting for lifelong person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2889–2897, 2021. 2, 6

[36] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv preprint arXiv:1604.01850*, 2(2):4, 2016. 5

[37] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):2872–2893, 2021. 1

[38] Chunlin Yu, Ye Shi, Zimo Liu, Shenghua Gao, and Jingya Wang. Lifelong person re-identification via knowledge refreshing and consolidation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3295–3303, 2023. 1, 2

[39] Binjie Zhang, Yixiao Ge, Yantao Shen, Shupeng Su, Fanzi Wu, Chun Yuan, Xuyuan Xu, Yexin Wang, and Ying Shan. Towards universal backward-compatible representation learning. *arXiv preprint arXiv:2203.01583*, 2022. 2, 3

[40] Bo Zhao, Shixiang Tang, Dapeng Chen, Hakan Bilen, and Rui Zhao. Continual representation learning for biometric identification. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1198–1208, 2021. 6, 7

[41] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1116–1124, 2015. 1, 5, 6

[42] Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016. 1

[43] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3754–3762, 2017. 1, 5

[44] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1318–1327, 2017. 1

[45] Da-Wei Zhou, Fu-Yun Wang, Han-Jia Ye, Liang Ma, Shiliang Pu, and De-Chuan Zhan. Forward compatible few-shot class-incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9046–9056, 2022. 3